

Multi-view Orientation Estimation using Bingham Mixture Models

Sebastian Riedel, Zoltan-Csaba Marton, Simon Kriegel

Institute of Robotics and Mechatronics, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany

Email: {firstname.lastname}@dlr.de

Abstract—This paper describes a multi-view pose estimation system, that is exploiting the mobility of a depth sensor through mounting it onto a robotic manipulator. Given a pose estimation algorithm that performs feature extraction and matching to a model database, we investigate the probabilistic modeling of the pose space as well as the measurement uncertainty, to be used in a sequential state estimation approach.

Uncertainties in 3d position can be modeled in a parametric way by 3d Gaussians, but the space of rotations in 3d - the special orthogonal group $SO(3)$ - requires approaches from directional statistics. A convenient representation for orientations are unit quaternions over which the Bingham distribution defines a parametric probability density function. The Bingham distribution also correctly accounts for the sign symmetry of orientation quaternions and leave degrees of freedom unconstrained (which is especially useful if an object is rotationally symmetric, with no unique quaternion describing its orientation).

In our experiments we test different sequential fusion methods, optimize their parameters, and investigate how the derived filter performs in a case with high uncertainties.

I. MOTIVATION

Recognition and pose estimation of known objects is necessary for many tasks including monitoring and tracking purposes or robotic manipulation of objects. Whereas recognition is the task of deciding which object is present, pose estimation refers to estimating an object's position and orientation in up to three dimensions. If the objects are known in advance, analyzed by the estimation algorithm in an offline training phase and the online application is limited to the a priori known objects, one speaks of model based object recognition and pose estimation. Robotic part handling in industrial applications is a prominent example and commercial use case for such algorithms as industrial manipulation is usually limited to a fixed set of parts known in advance.

Model based recognition and pose estimation have been subject to extensive research since the early 70s [1] and generally work in two steps. In the offline phase, a representation of the object is built using features derived from training data. In the online phase, incoming sensor data is matched to this representation and by doing so, the desired quantity - for example the object's pose - is measured. The quality of this measurement is affected through several aspects of the measurement process. Aspects involving the sensor directly include for example the inherent loss of 3d information when working with monocular 2d intensity images, loss of color information in 3d range data, limited spatial and temporal resolution, limited field of view and sensor noise. The sensed data is affected by its environment for example by scene illumination and occlusions. Lastly, the objects to be recognized and located can share

feature characteristics with the environment, among each other or among different views of the same object. This last aspect can lead to classification and pose ambiguities even under perfect environment and noise free sensing conditions.

Aforementioned influences are minimized by designing the extracted features to be more or less invariant to many of these aspects. This way, state-of-the-art algorithms achieve correct recognition results and pose measurements using only a single view of the scene or object. In case of ambiguities or inaccuracies introduced by the environment, the object itself or simply sensor noise, it can be helpful to acquire several sensor measurements and fuse the information obtained by them [2]. Utilizing multiple measurements in general will lead to higher accuracy of the measured quantities. Furthermore, tasks for which the presence of unknown objects is expected often necessitate fusing information from multiple measurements to obtain reliable estimates (for example in object search or scene exploration as explained in [3]). It is therefore purposive to employ approaches based on multiple views for scenarios, environments or objects where large inaccuracies are to be expected.

In this paper, we present an approach for estimation the object orientation applying multiple views using a Bingham Mixture Model (BMM).

II. SEQUENTIAL ORIENTATION ESTIMATION USING BINGHAM MIXTURE MODELS

This section reviews the Bingham distribution over unit quaternions and describes the formalism and two implementations for sequentially estimating an object's orientation using Bingham Mixture Models as probabilistic representation.

A. The Bingham Distribution over Unit Quaternions

The Bingham distribution is a distribution over n -dimensional hyperspheres $S^n, n \in \mathbb{N}^+$ [4]. The distribution is always antipodally symmetric, thus in the case of $n = 3$, it correctly captures the antipodal symmetry of the quaternion parametrization of rotations in 3d. The Bingham distribution over quaternions is derived from a 4d zero-mean Gaussian density

$$p(\mathbf{x}^*; C) = \frac{1}{\sqrt{(2\pi)^4 |C|}} \exp\left(-\frac{1}{2} \mathbf{x}^{*T} C^{-1} \mathbf{x}^*\right) \quad (1)$$

where $\mathbf{x}^* \in \mathbb{R}^4$ and C is a covariance matrix. This standard Gaussian is defined over all \mathbb{R}^4 and by re-normalizing it we can obtain a proper probability distribution over S^3 and thus 3d rotations. Intuitively, the process of renormalization

can be thought of as intersecting the unit quaternion sphere with the volumetric Gaussian density in 4d and normalizing the resulting density on the sphere to integrate to one. The Gaussian form of the Bingham distribution is thus

$$p(\mathbf{x}; C) = \frac{1}{F} \exp(\mathbf{x}^T C^{-1} \mathbf{x}) \quad (2)$$

with $\mathbf{x} \in S^3$ is now constrained to lie on the unit sphere and F is chosen so that

$$\int_{\mathbf{x} \in S^3} \exp(\mathbf{x}^T C^{-1} \mathbf{x}) = 1 \quad (3)$$

The standard form of the Bingham distribution is given by

$$\mathcal{B}(\mathbf{x}; \mathcal{K}, V) := \frac{1}{F(\kappa_1, \kappa_2, \kappa_3)} \exp\left(\sum_{i=1}^3 \kappa_i (v_i^T \mathbf{x})^2\right) \quad (4)$$

with $\mathcal{K} = (\kappa_1, \kappa_2, \kappa_3)^T$, $\kappa_i \in \mathbb{R}^- \cup \{0\}$ denoting the concentration parameters, $V = [v_1 | v_2 | v_3]$, $v_i \in S^3$ a set of orthogonal basis vectors and $\mathbf{x} \in S^3$ a unit quaternion. The standard form is based on an eigenvector/eigenvalue decomposition of the Gaussian covariance matrix C , zero-ing out $\hat{\kappa}_4$ and imposing an ordering constraint on the eigendecomposition. By choosing the eigenvectors $v_i, i \in \{1, 2, 3\}$ one adjusts the rotation axis in which rotational uncertainty applies. By choosing the concentration parameters, the kind of rotational uncertainty is specified. Two easy interpretable choices are:

- $\kappa_1 = \kappa_2 = \kappa_3 = \alpha \leq 0$: Gaussian-like rotational uncertainty. For $\alpha = 0$ the distribution is completely uniform. The more negative the α gets, the more peaked the distribution gets around its mode v_4 where v_4 is uniquely defined (orthonormal completion) by $\langle v_4, v_i \rangle = 0; i = 1, 2, 3$.
- $\kappa_1 = \kappa_2 \ll \kappa_3 \leq 0$: Increasing the value of κ_3 leads to increased uncertainty of the 4d-Gaussian in the principal component direction v_3 . Setting the concentration parameter κ_3 to zero results in an infinite variance along the direction v_3 for the 4d Gaussian and in terms of the normalized distribution over rotations this results in a density function mimicking a great circle on the unit quaternion sphere. When v_3 has the special form of $v_3 = (0, a, b, c)$ this axis around which the uniform rotation occurs is exactly (a, b, c) .

In Figure 1 the rotational uncertainty for different values of $\kappa_3 \in \{-900, -480, 0\}$ and $v_3 = (0, 0, 0, 1)$ is shown and the increase in uncertainty for rotations around $(0, 0, 1)$ is clearly visible. The plots shown are similar in nature to EGI (extended Gaussian image) plots and an intuitive way to visualize probability distributions over rotations. They work by rotating a base point on the sphere by many sampled rotations of the distribution. At the rotated point's location, a counter is increased. When visualizing the counts via a heat map, this gives a visualization of how the distribution behaves. In the plots in Figure 1 the point $p = (1, 1, 1)/\|(1, 1, 1)\|$ is rotated by $n = 100000$ sampled rotations. The mode of the distribution is the identity rotation, hence for $\kappa_3 < 0$ the peak is at p .

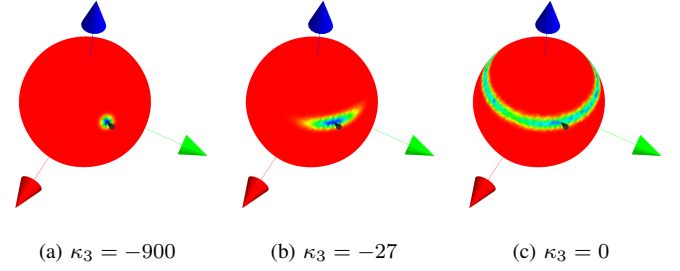


Fig. 1: EGI plots (see text for details) showing the effect of different concentration parameters κ_3 . Increasing κ_3 leads to increased rotational uncertainty around the z-axis.

For $\kappa_3 = 0$, we see the expected rotational invariance around the z-axis which is a consequence of choosing $v_3 = (0, 0, 0, 1)$.

The major problem with using the Bingham distribution in practice is the normalization constant $F(\kappa_1, \kappa_2, \kappa_3)$ which is expensive to compute [5]. Glover therefore precomputes the normalization constant for a discrete set of concentrations parameters ranging in $\kappa_i \in [-900, 0]$ and interpolates them as necessary.

B. Bingham Mixture Models

While individual Bingham distributions are in general unimodal (except one or more concentration parameters are zero), a sum of several Bingham distributions can represent arbitrary probability landscapes over rotations given a sufficient number of components. Formally, a Bingham Mixture Model (BMM) is defined as

$$\text{BMM}(\mathbf{x}; \{\alpha_i\}, \{\mathcal{K}_i\}, \{V_i\}) := \sum_i \alpha_i \mathcal{B}(\mathbf{x}; \mathcal{K}_i, V_i) \quad (5)$$

where $\mathcal{B}(\mathbf{x}; \mathcal{K}_i, V_i)$ is Bingham distributed as defined in equation (4) and $\sum_i \alpha_i = 1$. Unfortunately, BMMs - just like Gaussian mixture models - lose some of their analytical benefits as operations like maximum likelihood parameter fitting, entropy calculation and Kullback-Leibler (KL) divergence calculation are not defined analytically anymore and have to be approximated using for example Monte-Carlo integration techniques. In the context of this paper, we need to be able to fit BMMs to quaternion samples, multiply two BMMs and extract the maximum mode of a BMM. For fitting a BMM to a sample set of quaternions $\{q_i\}$, Glover [6] describes a greedy sample consensus method. The multiplication of two BMMs is done by pairwise multiplication of the Bingham distributions of both mixtures. As the result of multiplying two Bingham distributions is again a Bingham and can be done in an algebraic way, multiplication of two BMMs is also well-defined. Finding the maximum mode of a BMM cannot be done in closed form. An approximation is computed by sampling a high number of rotations from the BMM (typically 50000), evaluating their probability density value under the BMM and selecting the sample with the maximum value. This procedure will be used in later sections whenever the MAP estimate of a

BMM belief distribution is computed. For sampling a Bingham distribution, Glover [5] introduces an efficient Metropolis-Hastings sampler with the projected 4d-Gaussian density as proposal distribution.

C. Sequential Orientation Estimation using BMMs

The state estimation problem for sequentially estimating an object rotation using Bayes' theorem can be written as

$$p(q_t|z_t, \dots, z_0) \propto p(z_t|q_t)p(q_t|z_{t-1}, \dots, z_0) \quad (6)$$

where $q_t, z_t \in SO3$ are 3d rotations which are without loss of generality assumed to describe an object's rotation (frame o) in the world frame w ($q_t := {}^w q_{o,t}$ resp. $z_t := {}^w z_{o,t}$). Equation (6) does not contain a dynamic model as found in typical state estimation formulations as for our use case of rotation estimation of objects for industrial robotic manipulation we can assume static objects and precise robot motions. The measurement process underlying $p(z_t|q_t)$ can be described as producing a rotation measurements

$$z_t = w_t \circ q_t \quad (7)$$

based on the object's rotation q_t corrupted by a Bingham mixture distributed independent measurement noise $w_t \sim \text{BMM}(\alpha, \mathcal{K}, \mathbf{V})$ with \circ denoting quaternion multiplication. Based on findings of Glover [5], the conditional distribution $p(z_t|q_t)$ is Bingham mixture distributed according to $z_t|q_t \sim \text{BMM}(z_t; \alpha, \mathcal{K}, \mathbf{V} \circ q_t)$ where $\mathbf{V} \circ q_t$ are the eigenvectors of the original Bingham mixture components rotated by the fixed q_t . This mixture is a distribution over z_t given a fixed q_t . To apply this model in practice we can rewrite the distribution to obtain one over q_t given a fixed z_t by reordering the terms in the same way as presented in [5]

$$\text{BMM}(z_t; \alpha, \mathcal{K}, \mathbf{V} \circ q_t) = \quad (8)$$

$$\sum_{i=1}^N \alpha_i \left(\frac{1}{F_i} \exp \sum_{j=1}^3 \kappa_i((v_i \circ q_t)^T z_t)^2 \right) = \quad (9)$$

$$\sum_{i=1}^N \alpha_i \left(\frac{1}{F_i} \exp \sum_{j=1}^3 \kappa_i((v_i^{-1} \circ z_t)^T q_t)^2 \right) = \quad (10)$$

$$\text{BMM}(q_t; \alpha, \mathcal{K}, \mathbf{V}^{-1} \circ z_t) \quad (11)$$

If we now assume a Bingham mixture prior $p(q_t|z_{t-1}, \dots, z_0) = \text{BMM}(q_t; \beta, \mathcal{K}^*, \mathbf{V}^*)$, the posterior is also Bingham mixture distributed by multiplying the prior and measurement BMM. The components of the posterior BMM are built by pairwise multiplication of the prior and measurement mixture components

$$p(q_t|z_t, \dots, z_0) = \quad (12)$$

$$\sum_{i=1}^N \sum_{j=1}^M \alpha_i \beta_j \mathcal{B}(q_t; \mathcal{K}_i, V_i^{-1} \circ z_t) \mathcal{B}(q_t; \mathcal{K}_j^*, V_j^*) \quad (13)$$

As the multiplication of two Bingham is well-defined, Equation (13) describes a unique algebraic solution to this sequential estimation problem.

1) *Fusion based on Mixture Reduction*: In the algebraic posterior defined by Equation (13) we see that for a prior and measurement distribution with N and M components the resulting posterior mixture will have $N \times M$ components. To avoid a rapidly increasing number of components with every estimation step, mixture reduction strategies can be used to reduce the number of components while maintaining an acceptable level of representational accuracy with respect to the unreduced mixture. For the evaluation of our sequential Bingham mixture fusion, the reduction method of Runnalls [7], originally developed for Gaussian mixtures, has been implemented and tested for Bingham mixtures. It works by iteratively choosing two components of the mixture and merging them into one component. This is done until a specified number of components is reached. The two components to be merged are chosen by an upper bound criteria derived by Runnalls. It states that the the KL divergence of the mixture before the merge (\mathcal{M}_{i+j}) and after the merge (\mathcal{M}_{ij}) of components i and j is smaller than

$$d_{kl}(\mathcal{M}_{i+j}, \mathcal{M}_{ij}) \leq \alpha_i d_{kl}(B_i, B_{ij}) + \alpha_j d_{kl}(B_j, B_{ij}) \quad (14)$$

with B_i and B_j denoting the individual original components, B_{ij} denoting the merged component and α_i, α_j are the respective components' weights. As B_i, B_j and B_{ij} are single Gaussians (or Bingham), the KL divergence is analytically computable. Every iteration step, the two components leading to a minimal approximated before-after KL divergence are chosen to be merged. An example reduction is illustrated in Figure 2 by reducing an eight component mixture gradually down to one component.

The final algorithm for state fusion based on algebraic multiplication thus has two steps. At first, the full posterior is computed through element-wise multiplication as in equation (13). In the second step, the full posterior is reduced to a predefined maximum number of allowed components N_{\max} , which is the only parameter of this fusion method. For notational convenience, we denote this fusion algorithm as multiply & reduce (M+R) method.

2) *Fusion based on Sequential Monte-Carlo Sampling*: A different way to compute the posterior distribution was implemented using a standard sequential Monte Carlo (SMC) approach. In this approach, the posterior is represented by a set of M equally weighted samples $\{q_t^m\}, m = 1, \dots, M$ distributed according to the posterior distribution $\{q_t^m\} \sim p(q_t|z_0, \dots, z_t)$ [8]. The standard importance weighting and resampling steps are implemented by evaluation the measurement model measurement model $p(z_t|q_t)$ [8] and drawing samples from the weighted set with replacement and proportional to the weights. To allow the sample set to change "position", a Bingham mixture is fitted to the posterior sample set using the method described in [6] and sampled again in order to obtain the prior sample set for the next time step. The main parameter of the SMC method for state fusion is the number samples M . In all experiments later on $M = 100000$, which provides a good balance between computation time an representational accuracy.

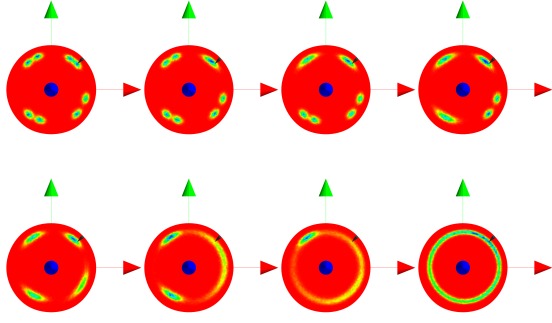


Fig. 2: EGI plots illustrating the implemented Bingham mixture reduction: The initial Bingham mixture (top, left) has eight components and roughly represents a 4-fold symmetric distribution around the z-axis (two components per fold). From left to right, top to bottom, the mixture gradually gets reduced by one component.

III. EXPERIMENTS

We evaluated the presented sequential BMM fusion methods over a range of parametrizations in two test scenarios dealing with the orientation estimation of real world objects for robotic grasping and manipulation.

In experiment III-A, test data from a previous experiment of our group has been reused [9]. The test data contains object pose measurement sequences over multiple views for three industrial objects. In this experiment, the obtained results using BMM fusion are compared to previous results from [9].

The second experiment, III-B, was chosen to evaluate orientation estimation under significantly higher uncertainty. It is based on a less sophisticated method for object orientation estimation which necessitates a more complex BMM measurement model. In this experiment, the theoretical minimum for estimating a unique object orientation due to two uncertainty in the measurement process is two sequence views.

Both experiments are based on real sensor data collected using a 6 DoF industrial robot, the Kuka KR16-2, with a mounted Asus Xtion as in [3].

A. Sequential Orientation Estimation under Moderate Uncertainty Conditions

For this experiment the test data contains three object pose measurement sequences corresponding to three industrial objects (*valve*, *filter* and *control*). The measured object poses were obtained using the object recognition and pose estimation method from [3] with an average rotational error of around four degrees. All sequences contain 20 views of their object from different viewing directions and all views lead to successful detection, hence 20 rotation measurements $z_t, t = 1, \dots, 20$ exist for every object. Whereas the *valve* and *control* object have a clear unique pose, the *filter* object has a strong rotational ambiguity in form of a 4-fold rotational symmetry. It is thus detected with a rotational error of $\pm 90^\circ$ or 180° in 7 out of the 20 views.

Two Bingham mixture measurement models have been defined for evaluation of the test sequences. The first one is the simple uninformed standard model where we assume the rotation measurement results from the correct pose corrupted by small Gaussian-like noise. A measurement model $p(z_t|q_t)$ describing this noise characteristic is created by rotating the correct pose by a Bingham with mode at identity and concentration parameters specifying a small Gaussian-like deviation, specifically

$$\mathcal{K} = \kappa \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \mathcal{V} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$p(z_t|q_t) = \mathcal{B}(z_t; \mathcal{K}, \mathcal{V} \circ q_t) \quad (16)$$

To deal with outliers, a uniform Bingham $\mathcal{B}_{uni}(z_t) = \mathcal{B}(z_t; (0, 0, 0), \mathcal{V})$ is added to the measurement model. This results in the Bingham mixture model

$$p_G(z_t|q_t) = \alpha_1 \mathcal{B}(z_t; \mathcal{K}, \mathcal{V} \circ q_t) + \alpha_2 \mathcal{B}_{uni}(z_t) \quad (17)$$

which can be "inverted" to a distribution over q_t by following equations (8) to (11). The Gaussian measurement model p_G has two parameters: the measurement concentration κ and the outlier ratio $\alpha = (\alpha_1, \alpha_2)$. This measurement model will be used for all three objects.

To test a more complex measurement model, the *filter* object was additionally evaluated with a measurement model taking the 4-fold symmetry around the object's z-axis into account. It is based on a five component Bingham mixture model. The first four components describe rotations of $\{0, +90, -90, 180\}$ degrees of the correct pose q_t around the symmetry axis and the fifth component again a uniform distribution. Construction of the symmetric measurement model p_S is a straightforward extension of equation (17) by adding the symmetric components. The resulting parameters for the model p_S are again the measurement concentration κ used for the first four components and the component weighting $\alpha = (\alpha_1, \dots, \alpha_5)$.

The metric for comparing different parameter settings for the measurement models $\{p_G, p_S\}$ and state fusion strategies $\{SMC, M+R\}$ is based on the maximum a posteriori (MAP) rotation error with respect to the ground-truth orientation averaged over the last ten views of every viewing sequence and all three viewing sequences.

For the Gaussian-like measurement model p_G the evaluated parameter range is listed in Table I. A summary of the top ten parametrizations is given in table II, which also includes the maximum MAP error within the last ten views over all three sequences and the percentages of MAP estimates below one, three and five degrees of error for the last ten views.

Out of the 72 parametrizations, 34 perform very close to the best one with a difference of only 0.21° in the avg. MAP error and 0.64° in the maximum MAP error. Seven of the top 34 runs are based on the SMC fusion, 27 are based on M+R. The outlier ratio α does not have an impact on

the obtainable performance, the measurement concentration κ does. All of the 34 good performing runs have a concentration $\kappa \in \{-27, -60, -120\}$, which correspond to 68% of the samples drawn from such a distribution having an rotational deviation smaller than $\{29^\circ, 19^\circ, 14^\circ\}$ from the mode of the distribution. This is somewhat contradicting to the actual measurement error in the object sequences, as for the *control*, *filter* and *valve* sequence 68% of the measurement errors lie below 4.4° , 6.9° and 4.4° . Thus even a concentration value of $\kappa = -240$ resp. 9.8° for the 68% interval encloses the measurement uncertainty well enough.

Figure 3 shows the MAP error evolution of the best SMC and M+R method compared to the error obtained by several other approaches to orientation fusion evaluated in [9] for the *filter* object sequence. As can be seen, the Bingham mixture approaches yield competitive results to a particle filter, a pose clustering and a histogram filter approach.

TABLE I: Explored parameter ranges for Gaussian-like measurement model evaluation

Functional Component	Parameter	Explored Range
SMC	M	{100000}
M+R	N_{\max}	{1, 5, 10}
p_G	α	{(.9, .1), (.95, .05), (.99, .01)}
	κ	{-900, -480, -240, -120, -60, -27}

TABLE II: Top 10 parametrizations (avg. MAP error) for the fusion evaluation of the Gaussian-like measurement model.

avg. MAP	max. MAP	% < 1°	Fusion	N_{\max}	α	κ
.57°	1.18°	.90	M+R	1	[.95, .05]	-60
.58°	1.16°	.90	SMC		[.90, .10]	-27
.59°	1.03°	.97	SMC		[.95, .05]	-60
.60°	1.42°	.90	M+R	1	[.90, .10]	-60
.61°	1.21°	.90	M+R	1	[.95, .05]	-27
.62°	1.30°	.87	M+R	1	[.90, .10]	-27
.62°	1.15°	.90	M+R	1	[.99, .01]	-60
.62°	1.31°	.90	M+R	5	[.95, .05]	-60
.62°	1.31°	.93	M+R	1	[.99, .01]	-27
.63°	1.36°	.90	M+R	10	[.99, .01]	-27
...						

The evaluation of the multimodal 4-fold symmetric measurement model p_S was carried out solely based on the test sequence of the *filter* object, as only this object shows this kind of measurement ambiguity (cf. the raw measurement line in the filter subplot of Figure 3). A total of 168 parametrizations has been evaluated exploring combinations of the two fusion methods, with both measurement models p_G and p_S and parametrizations given in table III.

Out of the 20 measurements in the filter sequence, seven are $\pm 90^\circ$ or 180° off, due to the 4-fold symmetry of the filter object. Six of these seven outliers fall into the last ten views of the sequence, over which the avg. MAP error metric is built. This evaluation thus puts special emphasis on how the parametrizations handle this ambiguity. The first 43 parametrizations yield similar performance in terms of average

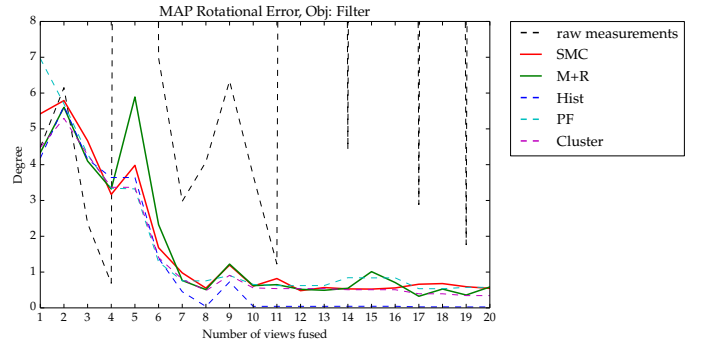


Fig. 3: MAP Rotational Error plots for the best SMC vs. the best M+R parameter settings in comparison to a histogram filter (discretized rotation space, Hist), a particle filter (PF) and a pose clustering approach (Cluster).

TABLE III: Explored parameter ranges for multimodal measurement model evaluation

Functional Component	Parameter	Explored Range
SMC	M	{100000}
M+R	N_{\max}	{1, 5, 10}
p_G	α	{(.9, .1), (.95, .05), (.99, .01)}
	κ	{-900, -480, -240, -120, -60, -27}
p_S	α	{(.25, .22, .22, .22, .1), (.45, .15, .15, .15, .1), (.65, .08, .08, .08, .1), (.85, .02, .02, .02, .1)}
	κ	{-900, -480, -240, -120, -60, -27}

MAP error (best/worst difference: 0.39°) and maximum MAP error (best/worst difference: 1.23°). The first 4-fold symmetric parametrization using a significant weight for the symmetric components ($\alpha_{2...4} = 0.15$) is ranked 50th with an avg./max. MAP error of $1.14^\circ/1.37^\circ$. The worst parametrization using p_S and the SMC fusion is ranked 85th with an avg./max. MAP error of $5.39^\circ/10.02^\circ$, while the vast majority of p_S -parametrizations with M+R fusion at some point return an MAP estimate around one of the outlier measurements and therefore obtain much worse MAP averages and maximum errors of up to $\approx 180^\circ$. It can be concluded, that both fusion methods perform better with the simple Gaussian measurement model and in general, if used with a complex measurement model, the SMC fusion has an advantage over the M+R fusion.

B. Sequential Orientation Estimation using BMMs under High Uncertainty Conditions

In this experiment, we estimate an object's orientation by means of a probabilistic viewing direction classification. Using a set of training images from 28 different viewing angles distributed across a half-sphere around the object, a classifier is trained in an offline processing step to predict the viewing direction from which an object feature can be observed from. During sequential estimation, the probabilistic prediction $p(D|f)$ over directions D given the feature f is used to build a measurement distribution $p(z_t|q_t)$. Classifiers have

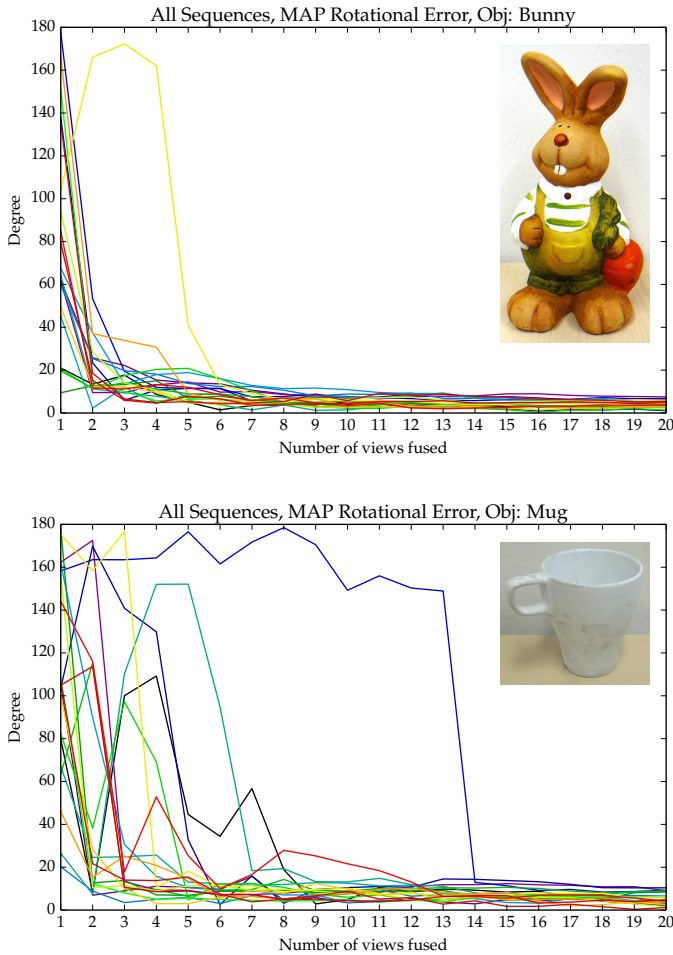


Fig. 4: MAP Rotational Error plots for 20 random sequences with 20 views.

been trained for two household objects: a decoration *bunny* object and a standard coffee *mug*. For evaluation, 20 view sequences showing each object from 20 random non-training viewing directions have been recorded.

To fuse this information over multiple views in this experiment, the measurement model has to account for two kinds of uncertainties: 1) the uncertainty in the viewing direction classification and 2) for every viewing direction the uniform orientation ambiguity around this viewing axis. The second uncertainty arises for example when using the popular Viewpoint Feature Histogram [10] or similar view-classification systems. We use a Bingham mixture model similar to the symmetric measurement model in experiment III-A to capture both uncertainties in a principled way. The measurement model contains 28 components, one component for every possible viewing direction. Every single component models the rotational invariance about its specific view axis by specifying v_3 appropriately, setting the corresponding κ_3 to 0 and the remaining κ -s to -27 for the *bunny* and -120 for the *mug* sequences.

In this experimental setup, using a only a single view of the

object means the orientation of it can only be estimated up to the rotational invariance around the camera optical axis (even if the viewing direction classification is 100% sure that we see it from a specific direction). For a unique orientation estimate, the sequential estimation requires at least two views from two non-parallel viewing directions. Based on the experience of the previous experiment, the SMC fusion implementation with $M = 100000$ samples has been used for this experiment.

Calculating the average MAP error over the last ten views as in experiment III-A (over all 20 sequences per object), the rotation estimation converges to an average error of 4.74° for the *bunny* and 9.08° for the *mug*. Figure 4 shows the MAP error evolution for every view sequence. While the resulting errors are too high to be of immediate practical use for robotic applications, the experiment showcases the successful information fusion of highly uncertain and ambiguous orientation estimates.

IV. CONCLUSION

We presented a sequential BMM-based orientation fusion for multi-view object pose estimation. Two variants of the method are evaluated for different parametrizations, and applied in two cases for orientation estimation of real world objects. We obtained competitive results to other approaches, and were able to significantly improve on the individual detections even on datasets where traditional methods would fail due to the high and non-Gaussian uncertainties.

V. ACKNOWLEDGMENTS

This work has partly been supported by the European Commission under contract number H2020-ICT-645403-ROBDREAM.

REFERENCES

- [1] R. T. Chin and C. R. Dyer, "Model-Based Recognition in Robot Vision," *ACM Computing Surveys (CSUR)*, vol. 18, no. 1, 1986.
- [2] S. D. Roy, S. Chaudhury, and S. Banerjee, "Active recognition through next view planning: a survey," *Pattern Recognition*, vol. 37, no. 3, pp. 429–446, 2004.
- [3] S. Kriegel, M. Brucker, Z.-C. Marton, T. Bodenmüller, and M. Suppa, "Combining Object Modeling and Recognition for Active Scene Exploration," *IEEE/RSJ IROS*, 2013.
- [4] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, 1974.
- [5] J. Glover and L. P. Kaelbling, "Tracking 3-D Rotations with the Quaternion Bingham Filter," MIT, Tech. Rep., 2013. [Online]. Available: <http://dspace.mit.edu/handle/1721.1/78248>
- [6] J. Glover, G. Bradski, and R. B. Rusu, "Monte Carlo pose estimation with quaternion kernels and the bingham distribution," *Robotics: Science and Systems*, 2012.
- [7] A. R. Runnalls, "Kullback-Leibler Approach to Gaussian Mixture Reduction," *IEEE TAES*, vol. 43, no. 3, pp. 989–999, Jul. 2007.
- [8] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
- [9] Z.-C. Marton and S. Türker, "On Bayesian Inference for Embodied Perception of Object Poses," in *IEEE/RSJ IROS: Workshop on Metrics of Embodied Learning Processes in Robots and Animals*, 2013. [Online]. Available: <http://elib.dlr.de/90326/1/bayesian.pdf>
- [10] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D recognition and pose using the Viewpoint Feature Histogram," *IEEE/RSJ IROS*, Oct. 2010.